

# DEEP LEARNING IN HEALTHCARE: HEART DISEASE DETECTION

\*Suchita Prakash Mandhare, \*\*Prof. Dr. Kamal Miyalal Alaskar

\*Assistant Professor, Somaiya School of Basic and Applied Sciences, Somaiya Vidyavihar University, Maharashtra, India  
Research Scholar, Department of Computer Applications,  
Bharati Vidyapeeth (Deemed to be University) Institute of Management Kolhapur, Maharashtra, India  
Email: [suchita.m@somaiya.edu](mailto:suchita.m@somaiya.edu)

\*\*Professor, Department of Computer Applications,  
Bharati Vidyapeeth (Deemed to be University) Institute of Management Kolhapur, Maharashtra, India  
Email: [kamal.alaskar@bharativedyapeeth.edu](mailto:kamal.alaskar@bharativedyapeeth.edu)

\*\*\*\*\*

## Abstract:

One of the most common and deadly diseases in the world is a Heart Disease. This Disease taking many lives every year. To preserve lives, this disease needs to be discovered early. One of the least, easiest, quickest and expensive ways to detect disease is using machine learning (ML) technique. Researchers are concentrating on developing intelligent systems that employ machine learning and deep learning algorithms for accurately identifying cardiac defects based on electronic health data because these disorders can be fatal. Using patient data on critical health indicators, the study presents unique machine learning algorithms to help forecast cardiac disease. The main goal of this study is to identify the patient who is more prone to develop heart problems based on a number of medical traits. to predict the chance of a heart disease, the model is created by utilizing the patient's medical record. To classify and anticipate Heart disease, different Machine learning algorithms were used like KNN, CNN, logistic regression. A very useful technique was used to limit the model's capacity to improve the accuracy of heart attack predictions for any individual. The suggested model demonstrated a satisfactory level of strength and demonstrated acceptable accuracy for forecasting heart disease in a given individual through the use of KNN and logistic regression.

**Keywords — Machine learning, Deep learning, Heart Disease, Healthcare, Convolution neural network (CNN), Medical image analysis, Supervised learning.**

\*\*\*\*\*

## I. INTRODUCTION

Providing healthcare is one of the main goals for humans. Everybody has the fundamental right to good health, in accordance with the principles of the World Health Organization. Regular health checks are believed to need access to suitable health care providers. Over 31% of fatalities worldwide are caused by heart-related disorders. It can be particularly tough to correctly identify and treat a range of cardiac issues due to a lack of diagnostic facilities, medical experts with sufficient

education, and other resources, especially in developing nations.[1]

By providing precise and effective disease recognition, machine learning techniques have the potential to drastically change the medical industry.

Heart diseases have been detected promptly, and a number of consequences have begun to become apparent [2] [3]. Reducing the risk of mortality is possible if illnesses associated with the heart can be identified earlier on. To find trends in the information and make inferences from it, several machine learning algorithms are

used on medical data. The majority of healthcare data have massive amounts and complex structures. Algorithms for machine learning have the ability to handle massive quantities of data and gain useful information from it. Historical data and real time input is used by Machine learning algorithms to make forecasts. Cardiologists may be encouraged to take action earlier by using this type of artificial intelligence framework for cardiac disease forecasting, which could preserve many lives by enabling more patients to obtain therapy in a shorter period of time. Prominent technologies that assist people working in any business determine unknown data's consequences from known facts are deep learning techniques. Numerous researchers used Machine learning models to forecast cardiac disease. Numerous common heart-related risk factors have been found, including raised cholesterol, diabetes, high bp, lack of physical activity, obesity which are considered controllable risk factors, and age, gender and genealogy, which are considered uncontrolled risk factors.

## II. RELATED WORK

The literature tackled heart disease prediction in several methods. The same inputs used in our study were employed in [4] to measure wearable smartphones and forecast the probability of cardiac illnesses with an accuracy of 84.6% using Naïve Bayes, SVM, and Functional Trees. Moreover, Naïve Bayes was utilized in [5] with the same dataset with a little higher accuracy of 86.5%.

The authors of [6] succeeded to achieve a 77.2% prediction accuracy through using the logistic regression approach on the Cleveland database, which is accessible at a UCI online data mining repository.

Data mining perspectives have been proposed by Chaurasia et al. [7] for recognizing cardiac problems in human bodies. Numerous data mining techniques, including J-48, Naïve Bayes, and bagging, have been used in this strategy, utilizing the WEKA tool for machine

learning. The machine learning portion involves the UCI laboratory. There are 313 features and 13 attributes for prediction in the heart disease dataset. In terms of classification accuracy, Naïve Bayes provides 82.30%, J48 provides 84.34%, and Bagging provides 85.34%.

R. Sharmila et al. 's proposed [8] to use data approaches to improve the dataset's prediction of cardiac diseases. SVM offered an accurate 85%. Sequential SVM is less accurate than parallel SVM with respect to accuracy.

A. K. Das, Chaudhuri, et al. [9] reported that no one strategy was effective. They compared numerous approaches such as random forest and logistic regression. K means clustering is implemented for validation, and its accuracy is 85.1%.

Using the Random Forest machine learning technique, Ricardo Buettner and Marc Schuster [10] introduced an innovative method to classify heart disease. A database gathering clinical data as well as patient test results was utilized to evaluate the approach's effectiveness. When the model used a 10-fold cross-validation approach, researchers noticed it exceeded other machine learning algorithms evaluated on the same dataset, with an overall accuracy of 84.45%. Additionally, the model showed an overall accuracy of 82.89% even in the absence of cross-validation.

## III. METHODOLOGY

### A. Data Collection

Kaggle was the source of the dataset [11]. Based on Table I's description, the dataset has 303 instances overall with 13 attributes.

Data element	Description
Age	
Sex	
Cp	Pain in the chest
Trestbps	Blood pressure at rest
Chol	Cholesterol value
Fbs	Value of fasting blood sugar

Restecg	Resting electrocardiographic results
Thalach	Maximum heart rate achieved
Exang	Exercise induced angina
Oldpeak	ST depression induced by exercise relative to rest
Slope	The peak exercise ST segment's slope
Ca	Number of major vessels
Thal	Defect type
Target	Diagnosis of heart disease

Table I DATASET

## B. Data Preprocessing

Data preparation is crucial because the caliber of the data that is used to build a model for machine learning significantly impacts its performance. Data preparation includes converting, resampling, and feature selection in addition to cleaning the data by removing outliers and corrupted or missing data points.

We began searching for any missing data, but we were not able to find any. Here are a few missing variables in the original dataset's characteristics, which may contribute to inaccurate outcomes and decrease the accuracy of the model. The "mean of column" approach is the most efficient method to replace the missing value in this situation. This method converts the mean or average of the surrounding numbers with 0 [12]. It then uses the right away estimated number to modify the 0 value. The values in the dataset were then transformed from numeric to nominal to make sure they worked with the machine learning methods that were being used. We searched for outliers and discovered a few, as Table II demonstrates.

Attributes	Outlier Values
Age of patient	None
Cholestrol	418,565,393,406,408

Trestbps	175,179,181,200,174,192
Thalach	71
Oldpeak	4.2,6.2,5.6,4.2,4.4

Table II LIST OF OUTLIERS

Only the severe outliers were eliminated since the mild outliers assist in establishing the final diagnosis.

## C. Building Model

The data is examined using various machine learning algorithms. If the method is supervised learning, then it will be applied to a training and testing set that has been constructed. After that, a cross-validation procedure is carried out to confirm that the earlier procedures were completed accurately.

The algorithm's accuracy will then be shown by the system when the cross-validation process is successful. The system will then stop when every process has been successfully performed. If the system determines after analysis that the data does not belong to a supervised learning method, it will immediately stop and begin analyzing the next data set.

In this work, a number of Machine learning methods such as decision tree, random forest, SVM were used. These techniques have been proven to be effective and produce positive outcomes, making them some of the best machine learning classifiers on the market.

### 1.1 Support Vector Machine (SVM)

For regression and anomaly detection, support vector machines are used. SVM are supervised learning technique. As seen in Figure 1, the goal is to label using one or more feature vectors with the goal to establish decision boundaries between various classes. The sites that are near to each other are referred to as support vectors.

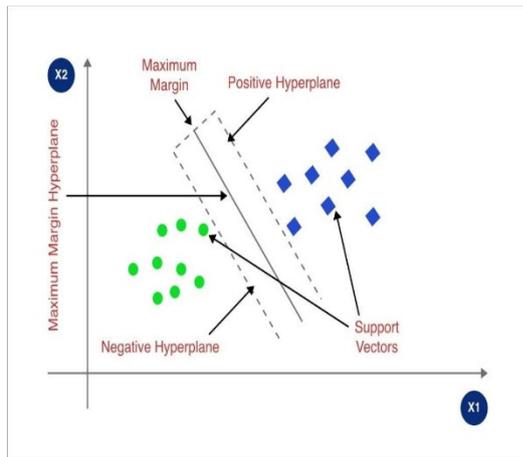


Fig1: Support Vector Machine

### 1.2 Decision Tree

Decision trees are a tool for visualizing algorithms. The classification models in machine learning are developed using the decision tree method [13]. The classification model's tree-like structure acts like its foundation. This comes inside the group of supervised learning, when the desired result is known in advance. The Decision Tree algorithm can be applied to both numerical and categorical data.

Multiple variables can triggered, including cholesterol levels, BP, weight, nicotine, and cholesterol. The most difficult component of the decision tree is choosing which node is the root. The element that is present in the root nodes needs to be used to categorize the data precisely. We utilize age as the root node. Understanding the decision tree is easy.

### 1.3 Random Forest Classification

When deployed with high-dimensional data, the Random Forest classifier provides great efficiency and accuracy in machine learning techniques. A group of unpruned classification-based trees is known as Random

Forest [14]. Because there is a small likelihood of overfitting and it is ineffective over noise in the dataset, it works incredibly well whenever applied to a variety of real-world applications. It typically increases the accuracy of data used for testing and validation. It also operates more quickly than other tree-based algorithms.

Similar to a decision tree, the Random Forest classifier fits a super defined number of decision trees on subsamples of the data rather than constructing a single tree for classification. This promotes both the classification power and the diversity of the system [16].

## IV. RESULT AND DISCUSSION

As the experiment was across, the results showed that Random Forest and SVM surpassed Decision trees. The SVM-developed model has a 98% accuracy rate, that is 8.01 percent more than and around 13.01 percent more than the decision tree. Similar to this, the Random Forest model surpasses our second-best SVM model with an accuracy of 99% for the prediction of heart attack., Regretfully, we were unable to locate a decision tree that worked effectively with the data that we have.

Model	Precision	Recall	ROC Area
SVM	0.994	0.994	0.994
Random Forest	0.996	0.996	1.00
Decision Tree	0.903	0.903	0.955

TABLE III. PERFORMANCE MEASURE OF MODEL

## V. CONCLUSION:

Decision Tree, Random Forest, and SVM are the modeling techniques utilized for machine learning that were evaluated in this study.

Based on the comparative analysis, SVM is the most suitable algorithm because it provides better overall results and only has a 2% lower recall score than the models that had the best recall value.

Additional effort is required to present additional data sets with the same properties in order to create an algorithm that is far more accurate. In addition, more features can be stated to the dataset to increase its diversity and potentially improve the accuracy of models with features that can be better correlated that yield better results.

## REFERENCES

- [1] J. Thomas and R. T. Princy. "Human heart disease prediction system using data mining techniques", 2016.
- [2] <https://www.who.int/hrh/links/en/>
- [3] D. Baroud, A. N. Hasan, and T. Shongwe. "A study towards implementing various artificial neural networks for signal classification and noise detection in PFDM/PLC Channels," 12th IEEE international symposium on communication systems, networks and digital signal processing (CSNDSP), 2020.
- [4] A. F. Otoom, E. E. Abdallah, Y. Kilani, A. Kefaye, and M. Ashour, "Effective diagnosis and monitoring of heart disease," *Int. J. Softw. Eng. its Appl.*, vol. 9, no. 1, pp. 143–156, 2015, doi: 10.14257/IJSEIA.2015.9.1.12.
- [5] K. Vembandasamp, R. R. Sasipriyap, and E. Deepap, "Heart Diseases Detection Using Naive Bayes Algorithm," *IJSET-International J. Innov. Sci. Eng. Technol.*, vol. 2, no. 9, 2015, Accessed: Dec. 11, 2021. [Online]. Available: [www.ijset.com](http://www.ijset.com).
- [6] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease," *The American journal of cardiology*, vol. 64, no. 5, pp. 304–310, 1989.
- [7] V. Chaurasia and S. Pal, "Data mining approach to detect heart diseases", *International Journal of Advanced Computer Science and Information Technology*, Vol.2, No.4, pp.56-66, 2014.
- [8] R. Sharmila, S. Chellammal, "A conceptual method to enhance the prediction of heart diseases using the data techniques", *International Journal of Computer Science and Engineering*, May 2018.
- [9] Chaudhuri, Avijit & Das, Anirban & Sinha, Deepankar & Banerjee, Dilip. (2021). Early Prediction of Heart Disease Using the Most Significant Features of Diabetes by Machine Learning Techniques. *ASIAN JOURNAL OF CONVERGENCE IN TECHNOLOGY*. 7. 10.33130/AJCT.2021v07i01.035.
- [10] Ricardo Buettner, Marc Schunter, "Efficient machine learning based detection of heart disease", 2019 IEEE International Conference on Ehealth Networking, Application & Services (HealthCom).
- [11] "Heart Disease UCI | Kaggle." <https://www.kaggle.com/ronitf/heart-disease-uci> (accessed Jan. 10, 2022).
- [12] <https://www.cs.waikato.ac.nz/ml/weka/mooc/dataminingwithweka/slides/Class5-DataMiningWithWeka-2013.pdf>
- [13] [https://en.wikipedia.org/wiki/Decision\\_tree\\_learning](https://en.wikipedia.org/wiki/Decision_tree_learning)
- [14] <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>
- [15] M. Liu, X. Xu, Y. Tao and X. Wan, "An Improved Random Forest Method Base on RELIEFF for Medical Diagnosis," 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference, pp. 44-49, 2017
- [16] Arooj S, Rehman SU, Imran A, Almuhaimeed A, Alzahrani AK, Alzahrani A. A Deep Convolutional Neural Network for the Early Detection of Heart Disease. *Biomedicine*. 2022 Nov 3;10(11):2796. doi: 10.3390/biomedicine1011279621. PMID: 36359317; PMCID: PMC9687844.
- [17] Tribikram Dhar, Nilanjan Dey, Senior Member, IEEE, Surekha Borra, Senior Member, IEEE, and R. Simon Sherratt, Fellow, IEEE Challenges of Deep Learning in Medical Image Analysis—Improving Explainability and Trust. *IEEE TRANSACTIONS ON TECHNOLOGY AND SOCIETY*, VOL. 4, NO. 1, MARCH 2023
- [18] D. -E. -M. Nisar, R. Amin, N. -U. -H. Shah, M. A. A. Ghamdi, S. H. Almotiri and M. Alruily, "Healthcare Techniques Through Deep Learning: Issues, Challenges and Opportunities," in *IEEE Access*, vol. 9, pp. 98523-98541, 2021, doi:10.1109/ACCESS.2021.3095312.
- [19] Jan Egger, Christina Gsaxner, Antonio Pepe, Kelsey L. Pomykala, Frederic Jonske, Manuel Kurz, Jianning Li, Jens Kleesiek, Medical deep learning—A systematic meta-review, *Computer Methods and Programs in Biomedicine*, Volume 221, 2022, 106874, ISSN 0169-

- 2607  
(<https://www.sciencedirect.com/science/article/pii/S1566253520303651>)
- [20] V. Marakala, G. V. Sriramakrishnan, G. Jakka, C. J. Shingadiya, H. P. Widiastuti and G. G. Reivan Ortiz, "Use of Deep Learning Application in Medical Devices," 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2022, pp. 935-939, doi: 10.1109/ICIRCA54612.2022.9985537.
- [21] Sowmya, H.N., Ajitha, S. (2022). A Study on Deep Learning Predictive Models in Healthcare. In: Joshi, A., Mahmud, M., Ragel, R.G., Thakur, N.V. (eds) Information and Communication Technology for Competitive Strategies (ICTCS 2020). Lecture Notes in Networks and Systems, vol 191. Springer, Singapore. [https://doi.org/10.1007/978-981-16-0739-4\\_81](https://doi.org/10.1007/978-981-16-0739-4_81)
- [22] K. Shailaja, B. Seetharamulu and M. A. Jabbar, "Machine Learning in Healthcare: A Review," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2018, pp. 910-914, doi:10.1109/ICECA.2018.8474918.
- [23] Kothari, V., Stevens, R.J., Adler, A.I., Stratton, I.M., Manley, S.E., Neil, H.A., Holman, R.R.: UKPDS 60: risk of stroke in type 2 diabetes estimated by the uk prospective diabetes study risk engine. *Stroke* 33(7), 1776–1781 (2002)
- [24] A. S. Cheema, S. Srivastava, P. Srivastava, and B. Murthy. A standard compliant blood bank management system with enforcing mechanism. In *Computing, Communication and Security (ICCCS)*, 2015 International Conference on, pages 1–7. IEEE, 2015.
- [25] X. Zheng, J. Shi, Y. Li, X. Liu, and Q. Zhang. Multi-modality stacked deep polynomial network based feature learning for alzheimer's disease diagnosis. In *Biomedical Imaging (ISBI)*, 2016 IEEE 13th International Symposium on, pages 851–854. IEEE, 2016.
- [26] Abdelhafiz, D., Nabavi, S., Ammar, R., Yang, C., Bi, J. 2019. Residual deep learning system for mass segmentation and classification in mammography. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB 2019)*, 475–484.
- [27] Al-A, Valliani A., Ranti, D., Oermann, E.K., 2019. Deep learning and neurology: A systematic review. *Neurol. Therapy* 8 (2), 351–365. <https://doi.org/10.1007/s40120-019-00153-8>
- [28] Francesco Piccialli, Vittorio Di Somma, Fabio Giampaolo, Salvatore Cuomo, Giancarlo Fortino, "A survey on deep learning in medicine: Why, how and when?", *Information Fusion*, Volume 66, 2021, Pages 111-137, ISSN 1566-2535, <https://doi.org/10.1016/j.inffus.2020.09.006>.
- [29] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. Snead, I. A. Cree, and N. M. Rajpoot. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE transactions on medical imaging*, 35(5):1196–1206, 2016.
- [30] R. Ranganath, A. Perotte, N. Elhadad, and D. Blei. Deep survival analysis. In *Machine Learning for Healthcare Conference*, pages 101–114, 2016.
- [31] Tsang, G., Xie, X., Zhou, S.-M. 2020. Harnessing the power of machine learning in dementia informatics research: Issues, opportunities, and challenges, *IEEE Rev. Biomed. Eng.* 13, art.no. 8665908, 113–129. <https://dx.doi.org/10.1109/RBME.2019.2904488>.
- [32] M. M. Uddin, K. Arfatul Islam, M. Mamun, V. K. Tiwari and J. Park, "A Comparative Analysis of Machine Learning-Based Website Phishing Detection Using URL Information," 2022 5th International Conference on Pattern Recognition and Artificial Intelligence (PRAI), Chengdu, China, 2022, pp. 220-224, doi: 10.1109/PRAI55851.2022.9904055.
- [33] K. Vembandasamp, R. R. Sasipriyap, and E. Deepap, "Heart Diseases Detection Using Naive Bayes Algorithm," *IJISSET-International J. Innov. Sci. Eng. Technol.*, vol. 2, no. 9, 2015, Accessed: Dec. 11, 2021. [Online]. Available: [www.ijisnet.com](http://www.ijisnet.com).
- [34] Saleem, M., Khadim, A., Fatima, M., Khan, M.A., Nair, H.K. and Asif, M., 2022, October. ASSMA-SLM: Autonomous System for Smart Motor-Vehicles integrating Artificial and Soft Learning Mechanisms. In *2022 International Conference on Cyber Resilience (ICCR)* (pp. 1-6). IEEE.
- [35] Ahmed, Usama, et al. "Prediction of diabetes empowered with fused machine learning." *IEEE Access* 10 (2022): 8529-8538.